

Estimating the sensitivity of centrality measures w.r.t. measurement errors

Christoph Martin Peter Niemeyer

Leuphana University of Lüneburg
Scharnhorststr. 1, 21335 Lüneburg, Germany
{cmartin,niemeyer}@uni.leuphana.de

March 31, 2017

Abstract

Most network studies rely on some observed network that differs from the underlying network which is obfuscated by sampling errors. It is well known that such sampling errors impact network metrics, e.g.: centrality measures. That is: a more central node in the observed network might be less central in the underlying network.

In order to measure this sensitivity of centrality measures with respect to sampling errors we study the probability, that given two randomly chosen nodes, the more central one in the observed network is also more central in the underlying network.

As this concept relies on the underlying network which is generally not accessible, we present first thoughts on how to approximate this sensitivity, given (i) the observed network and (ii) some assumptions on the underlying error-mechanism (e.g.: the percentage of missing edges or nodes).

Contents

1	Introduction	2
2	Basic concepts	3
2.1	Modeling measurement errors	4
2.2	Sensitivity of centrality measures	6
2.3	Exemplary application of the sensitivity concept	7
3	How to the estimate the sensitivity of a centrality measure	9
3.1	Methods to estimate the sensitivity	9
3.2	Example for method application	10

3.3	Evaluating estimation results	11
3.4	Results for synthetic graphs	13
4	Application to real-world networks	16
4.1	Results for real-world networks	17
5	Discussion and conclusion	20

1 Introduction

Most network studies rely on some observed network that differs from the underlying network which is obfuscated by sampling errors.

It is well known that such sampling errors impact network metrics, e.g.: centrality measures. That is: a more central node in the observed network might be less central in the underlying network.

In order to measure this sensitivity of centrality measures with respect to sampling errors we study the probability, that given two randomly chosen nodes, the more central one in the observed network is also more central in the underlying network.

As this concept relies on the underlying network which is generally not accessible, we present first thoughts on how to approximate this sensitivity, given (i) the observed network and (ii) some assumptions on the underlying error-mechanism (e.g.: the percentage of missing edges or nodes). To do so we follow two different approaches, which we call "imputation-based method" and "iterative method". In the first case we try to simulate the hidden network by inverting the error mechanism (e.g.: if the error mechanism is "10% of randomly chosen edges are missing", we add 11,11% edges by linking randomly chosen pairs of non-adjacent nodes), and then compute the sensitivity with respect to the simulated hidden network. In the iteration approach, we assume, that the considered hidden network has a certain kind of self-similarity property. That is, we assume that the sensitivity of the observed network (with respect to the given error-mechanism) is a good estimate for the sensitivity of the hidden network.

While both approaches turn out to work very well in the case of homogeneous networks (in our experiments we consider ER-random graphs), to our surprise the iteration approach outperforms the imputation approach in the case of inhomogeneous networks (BA random graphs in our experiment). We observe the same effect when applying both methods to a set real-world graphs.

In the past years, researcher have studied the reliability of centrality measures under the influence of measurement errors. As common sense suggests, centrality measures become less reliable with increasing level of error. However, the relationship between error level and reliability is highly dependent on the type of measurement error, the centrality measure, and the network

structure [Costenbader and Valente, 2003, Borgatti et al., 2006, Frantz et al., 2009, Wang et al., 2012, Smith and Moody, 2013]. For an extensive survey of previous studies about the reliability of centrality measures see [Smith et al., 2017].

Since errors in network data are omnipresent, much effort has been made to infer statistics about error-free network based on the erroneous network data or to recover the error-free network [Butts, 2003, Huisman, 2009, Kim and Leskovec, 2011, Frantz and Carley, 2016, Newman, 2017].

Most studies about the reliability of centrality measures mentioned above analyze the impact of measurement errors by manipulating error-free networks. In this study, we propose two methods that estimate the impact of measurement errors on the reliability of centrality measures but without using the error-free network. Thus researcher can apply these methods to their observed (erroneous) networks and incorporate the impact of the measurement error into their analyses. To measure the reliability of centrality measures, we introduce the necessary concepts in Section 2. The estimation methods are presented in Section 3. In Section 4, we apply these methods to real-world networks. Results are discussed in Section 5.

2 Basic concepts

Let G be an undirected, unweighted, finite graph with vertex set $V(G)$ and edge set $E(G)$.¹ A centrality measure c is a real-valued function that assigns centrality values to all nodes in a graph. Furthermore, a centrality measure is invariant to structure-preserving mappings, centrality values depend solely on the structure of a graph. External information (e.g., node or edge attributes) have no influence on the centrality values [Koschützki et al., 2005]. We denote the centrality value for node $u \in V(G)$ by $c_G(u)$, and the centrality values for all nodes in G by $c(G)$.

The following centrality measures are used in this study: closeness centrality, betweenness centrality [Freeman, 1978], degree centrality, eigenvector centrality [Bonacich, 1987], and the pagerank [Brin and Page, 1998]. All centrality measures are calculated using the igraph library (version 0.7.1) [Csardi and Nepusz, 2006].

Let G and G' be two graphs and c a centrality measure. There are $\binom{|V(G) \cap V(G')|}{2}$ possible ways to create pairs of nodes that are in G and G' . A pair of nodes $\{u, v\}$ with $u, v \in V(G) \cap V(G')$ and $u \neq v$ is called concordant w.r.t. c , if the order of u and v is the same in $c(G)$ as in $c(G')$, i.e. either $c_G(u) < c_G(v)$ and $c_{G'}(u) < c_{G'}(v)$ or $c_G(u) > c_G(v)$ and $c_{G'}(u) > c_{G'}(v)$. A pair of nodes is called discordant, if the order of u and v in $c(G)$ differs from the order of u and v in $c(G')$, i.e. either $c_G(u) < c_G(v)$ and $c_{G'}(u) > c_{G'}(v)$

¹In this study, we only consider unweighted and undirected graphs. However, most concepts introduced in this study should be easily transferable to directed graphs (and with some adjustments to weighted graphs).

or $c_G(u) > c_G(v)$ and $c_{G'}(u) < c_{G'}(v)$. Ties are neither concordant nor discordant.

A random graph consists of a finite set of graphs Ω equipped with a function P that assigns a probability to every graph in this set [Bollobás and Riordan, 2002].

2.1 Modeling measurement errors

Network data can be influenced by a variety of different measurement errors. Recently, [Wang et al., 2012] categorized measurement errors into six groups: false negative nodes and edges, false positive nodes and edges, and false aggregation and disaggregation. For example, when 10% of the edges are missing in the observed network data, the graph made from this observed data suffers from false negative edges.

In this study, we use error mechanisms to model measurement errors. An error mechanism, denoted by φ , is defined by a procedure that describes measurement errors that may occur during the data collection. For a graph G , $\varphi(G)$ is a random graph which probability distribution for the possible graphs depends on the error procedure that describes φ . Hence, all graphs that could be observed when the error procedure influences the data collections are possible outcomes of this random graph.

To illustrate the concept of an error mechanism consider the graphs illustrated in Figure 1. The initial graph is denoted by H (drawn in the upper left corner). We assume, that we know the error mechanism that compromises the data collection. For this example, we assume that the error mechanism φ is edges missing uniformly at random with an error level of 50%. All graphs in the set of possible outcomes for this random graph $\Omega = \{G_1, G_2, \dots, G_6\}$ are also shown in Figure 1. For this example, the probability function is $P(G_i) = \frac{1}{6}$. In this example, graphs in Ω occur with the same probability. However, this concept is not limited to uniform distributions.

In this paper, we study the following error mechanisms:

- Nodes missing uniformly at random (rm nodes): A fraction of nodes (and all edges connected to these nodes) is missing in the observed graph. All nodes have the same probability to be missing in the observed graph.
- Edges missing uniformly at random (rm edges unif.): A fraction of edges is missing in the observed network. All edges have the same probability to be missing in the observed graph.
- Edges missing proportional (rm edges prop.): A fraction of edges is missing in the observed network. The probability that an edge is missing in the observed network is proportional to the sum of the degree values of the endpoints.

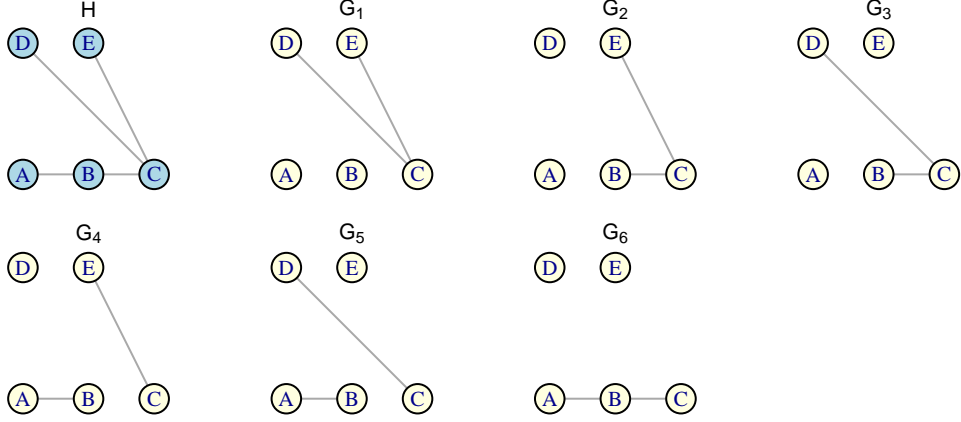


Figure 1: $\varphi(H)$ is a random graph with possible outcomes $\Omega = \{G_1, G_2, \dots, G_6\}$ and $P(G_i) = \frac{1}{6}$.

- Spurious edges (add edges): The observed graph contains too many edges. Every non-existing edge has the same probability to be erroneously observed.

Erroneous network data is ubiquitous and imputation techniques are commonly used to reconstruct the error-free network from noisy data [Huisman, 2009, Wang et al., 2016].

An imputation mechanism is a procedure that is applied to a network to ‘undo’ the effect of an error mechanism. We denote imputation mechanisms by ψ . Analogously to error mechanisms, $\psi(G)$ is a random graph. The possible outcomes of this random graph are all graphs that can occur if the imputation procedure ψ is applied to G . The procedure for an imputation mechanisms depends on the type of measurement error that has affected the data collection. Usually, multiple procedures are possible and choosing the correct one might be difficult (cf. Example 1).

In this paper, we use two imputation procedures: If k edges are missing in the observed network, we choose k edges uniformly at random from the set of edges that are not in the observed network and add them to the observed network. If there are k spurious edges in the observed network, we choose k edges that do exist in the observed network and delete them.

Example 1

For example, if edges are missing uniformly at random, a possible imputation procedure is to add edges randomly (uniform) to the observed network. However, the success of this procedure depends on the structure of the network. An (extreme) example: If the hidden network is, in fact, a star graph, it is quite likely that this imputation approach would have a negative impact on subsequent analyses because it is much more likely

that the imputed edges are between leaf nodes than between the internal node and a leaf node. Hence, if the structure of the hidden network is not known (which is usually the case), it is difficult to define an appropriate imputation procedure.

2.2 Sensitivity of centrality measures

Let G and G' be graphs on the same vertex set and c a centrality measure. The sensitivity of the centrality measure c w.r.t. those two graphs, G and G' , is the probability that two nodes with distinct centrality values, randomly chosen from the vertex set of G and G' , have the same order in $c(G)$ and $c(G')$, i.e., they are concordant. If this quantity is close to one, the centrality measure is considered to be robust. If it is close to zero, the centrality measure is considered to be sensitive. We calculate the sensitivity ρ for a centrality measure c with respect to G and G' as follows:

$$\rho_c(G, G') = \frac{n_c}{n_c + n_d} \quad (1)$$

With n_c as the number of concordant pairs and n_d as the number of discordant pairs w.r.t. the order given by $c(G)$ and $c(G')$.²

The sensitivity as defined above is closely related to Goodman and Kruskal's rank correlation coefficient γ which is the difference between concordant and discordant pairs divided by the sum of concordant and discordant pairs [Goodman and Kruskal, 1954]. We can use this relationship to calculate the sensitivity as follows:

$$\rho_c(G, G') = \frac{\gamma(c(G), c(G')) + 1}{2}. \quad (2)$$

Let us apply the concept of the sensitivity of centrality measures to the example from the previous section. Assume that we have observed the graph labeled as G_6 and that we are interested in the sensitivity of the degree centrality. We call the observed graph O . Then the degree centrality values are $\deg(H) = (1, 2, 3, 1, 1)$ for the nodes in H and $\deg(O) = (1, 2, 1, 0, 0)$ for nodes in O . Based on the degree values, we can calculate the sensitivity of the degree centrality with respect to O and H for our example: $\rho_{deg}(O, H) = \frac{5}{6}$. If we randomly choose a pair of nodes, there is a 0.83 chance that their order induced by the degree centrality is the same in both graphs.

² It may occur that $V(G') \neq V(G)$. In these cases, we only consider entries in $c(G)$ and $c(G')$ that correspond to nodes that are in both graphs (G and G'). This is a common approach for the comparison of graphs on different vertex sets. [leskovec]

2.3 Exemplary application of the sensitivity concept

In this section, we show how the concepts introduced above can be used to analyze the sensitivity of centrality measures in Erdős–Rényi graphs (ER graphs).

For all combinations of centrality measures and error mechanisms introduced in Section 2 and 2.1, we perform the experiment described below 500 times. For every error mechanism, we consider two cases, a moderate scenario of 10% error level and a more intense scenario with 30% error level. The procedure for the experiment is as follows:

1. Generate an ER graph with 100 nodes and edge probability 0.2 and denote it by H . This is the error-free (hidden) graph which not available to the researcher.
2. Choose a graph from $\varphi(H)$ and denote it by O . This is the observed graph which is affected by measurement errors.
3. Calculate the sensitivity $\rho_c(O, H)$.

The results are shown in Figure 2. Every panel shows violin plots for the distribution of the sensitivity of the centrality measures. Despite the fact that these graphs are very homogeneous, we make interesting observations. The sensitivity differs between centrality measures. The degree centrality is, in all cases, the most robust measure. Generally, the sensitivity values for the 30% error mechanisms are much lower than for the 10% error mechanism. The variance of the sensitivity values also increases with increasing error level. Usually, the sensitivity also depends on the error mechanism. However, this is not always the case. (e.g., degree centrality combined with 10% error mechanisms). These observations are conclusive with the results of [Borgatti et al., 2006]. These results show, that even for homogeneous networks, such as ER graphs, measurement errors have severe consequences for the reliability of centrality measures.

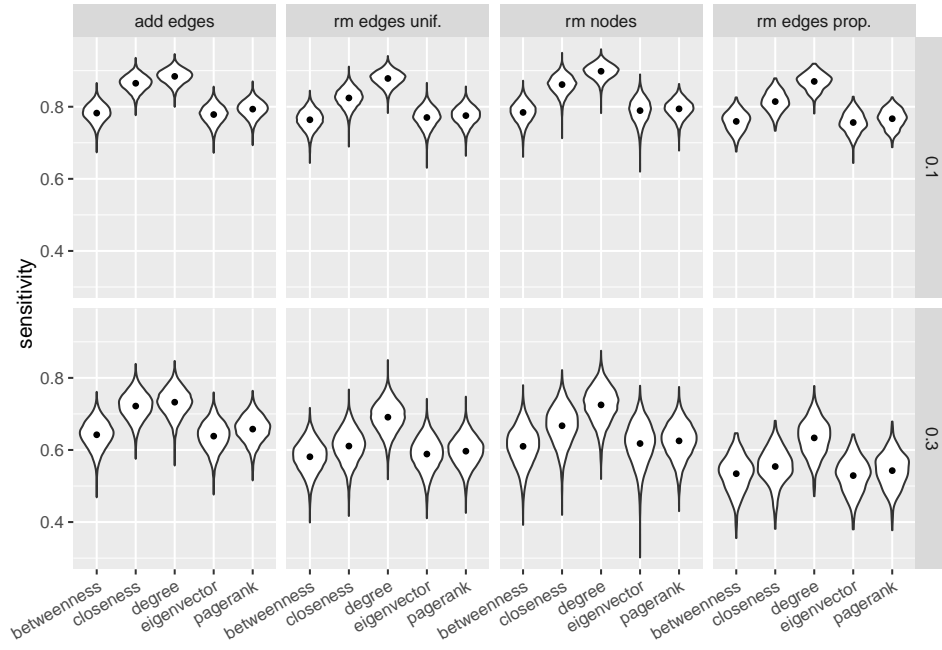


Figure 2: Sensitivity of centrality measures in Erdős-Rényi graphs. The violin plots in each show the distribution of the sensitivity of the corresponding centrality measure under the influence of different error mechanisms.

3 How to estimate the sensitivity of a centrality measure

Most network data is not error-free. The network that is derived from imperfect network data, the observed network, differs from the 'hidden' network that could be created from perfect data. Depending on the domain, subject-matter experts usually have some assumptions about the kind of measurement error that has affected the data collection. If a researcher or practitioner uses centrality measures to analyze the observed network, he wants to know how reliable the centrality values, derived from the erroneous network data, are.

Based on the notation introduced in Section 2, we can now formulate what information is available to the researcher and what he wants to know. After the data collection, we have the observed network O and we have assumptions about the error mechanism that has influenced the data collection. This error mechanism is denoted by φ . In addition to that, we denote the centrality measure of interest by c and the hidden network, the network that could be created from error-free data, is by H . Given this notation, we can formalize what the researcher wants to know: the 'true' sensitivity $\rho_c(O, H)$. Obviously, we can't calculate this expression since we don't have access to H .

3.1 Methods to estimate the sensitivity

Our first proposal for an estimation method for the sensitivity is imputation based. Based on the observed network O and an imputation mechanism, we try to 'reconstruct' the hidden network from the observed network. Based on the reconstructed network and the observed network, we calculate the imputation based estimate for the true sensitivity of the hidden network H and a centrality measure c as follows:

$$\hat{\rho}_c^{imp}(O, H) := E(\rho_c(O, \psi(O))). \quad (3)$$

We call this approach the imputation estimate for the true sensitivity. Since the $\psi(O)$ is a random graph, $\rho_c(O, \psi(O))$ is a random variable. To take the variability between different realizations into account, we use the expected value of this expression.

The actual form of the imputation mechanism ψ depends strongly on the error mechanism φ which has influenced the data collection. However, for some error mechanisms it might be very difficult to define an appropriate imputation mechanism (e.g., if nodes are missing depending on some external attributes). Therefore, we propose a second method to estimate the true sensitivity which does not depend on an imputation procedure. If we assume that the observed and the hidden network are similar w.r.t. the

sensitivity of a given centrality measure, we can estimate the true sensitivity as follows:

$$\hat{\rho}_c^{iter}(O, H) := E(\rho_c(O, \varphi(O))). \quad (4)$$

We call this approach the iterative estimate for the true sensitivity. The advantage of this approach is obvious: given the error mechanism φ , calculating the estimate is straightforward.

3.2 Example for method application

As a first step to check whether the proposed methods yield useful results, we use them to estimate the sensitivity of centrality measures in Erdős–Rényi random graphs under different error scenarios.

For all combinations of centrality measures and error mechanisms introduced in Section 2 and 2.1, we perform the experiment described below 500 times. For every error mechanism, we consider two cases, a moderate scenario of 10% error level and a more intense scenario with 30% error level. We perform the following steps to simulate erroneous data collection and to collect 3-tuples of true sensitivity, the estimate based on the iterative method, and the estimate based on the imputation method:

1. We generate an ER graph with 100 nodes and edge probability 0.2 and denote it by H . This graph represents the (error-free) hidden network.
2. We choose a graph from $\varphi(H)$ and denote it by O . This graph represents the observed network which is affected by measurement errors. For evaluation purposes, the true sensitivity $\rho_c(H, O)$ is calculated and denoted by s .
3. Based on the observed network O , two estimates for the true sensitivity are calculated. The imputation based estimate ($\hat{\rho}_c^{imp}(O, H)$) is denoted by \hat{s}^{imp} , the estimate calculated according to the iterative method ($\hat{\rho}_c^{iter}(O, H)$) is denoted by \hat{s}^{iter} .

The results of these experiments are listed in Table 1. Values in the s columns represent the mean values of the true sensitivity for all 500 runs. The 95th percentile values of the absolute difference between the true sensitivity and the estimate are labeled with \hat{e}^{imp} for the imputation based estimates and \hat{e}^{iter} for the estimates by the iterative method. We call this value the absolute error. For example, the average true sensitivity of the betweenness centrality under the influence of the error mechanism add edges random 10% is 0.891 and the imputation-based estimate of the true sensitivity is in the interval $[0.871, 0.911]$ in 95% of all runs.

It can be seen from Table 1 that there is a wide range of absolute error values (ranging from 0.017 to .059). How should these values be interpreted? For example, the estimates for the sensitivity of the degree centrality and

Table 1: Results for point estimate of sensitivity (ER graphs). For all centrality measures, average true sensitivity (column s) and 95th percentile of the absolute error for the point estimate (\hat{e}^{imp} : imputation based; \hat{e}^{iter} : iterative approach) are shown. (Values are multiplied by 100 for better readability.)

		bc				cc				dc				ec				pr			
		s	\hat{e}^{imp}	\hat{e}^{iter}		s	\hat{e}^{imp}	\hat{e}^{iter}		s	\hat{e}^{imp}	\hat{e}^{iter}		s	\hat{e}^{imp}	\hat{e}^{iter}		s	\hat{e}^{imp}	\hat{e}^{iter}	
error_type	lv																				
add edges	0.1	89.1	2.0	1.9	93.3	2.4	2.2	94.2	1.8	1.8	89.0	2.0	2.0	89.7	1.9	1.7					
rm e. prop		88.0	2.1	2.1	90.9	2.6	3.2	93.6	1.9	1.8	87.9	2.0	2.0	88.5	1.9	1.9					
rm e. unif		88.2	2.3	2.2	91.2	2.6	2.7	93.9	1.8	2.0	88.5	2.0	2.0	88.8	2.1	1.9					
rm nodes		89.2	–	2.3	93.1	–	2.4	95.0	–	1.8	89.6	–	2.4	89.7	–	2.0					
add edges	0.3	82.1	3.0	3.0	86.1	3.7	3.5	86.6	3.4	3.1	81.9	3.5	3.4	82.9	3.0	3.1					
rm e. prop		77.0	5.1	4.8	77.9	5.8	5.9	81.9	5.5	4.4	76.7	5.1	5.0	77.4	5.1	4.6					
rm e. unif		79.2	3.7	3.7	80.6	4.2	3.8	84.6	3.6	4.3	79.5	3.7	3.7	80.0	3.5	3.5					
rm nodes		80.5	–	4.3	83.3	–	5.0	86.2	–	4.6	80.9	–	5.0	81.2	–	4.2					

the pagerank under the influence of the add edges 10% error mechanism show approximately the same error values (ranging from 0.017 to 0.019). However, we argue that the estimate for the pagerank works better because the (average) sensitivity of the pagerank is 0.897 while the average sensitivity of the degree centrality is substantially higher (0.942). Therefore, one has to consider the magnitude of the true sensitivity when interpreting the absolute error of the estimate.

3.3 Evaluating estimation results

The estimate for the true sensitivity is successful if it is 'close' to the true sensitivity value. To determine this 'closeness', we calculate the error (absolute difference between true value and estimate) relative to the magnitude of the true sensitivity. On the one hand, if the true sensitivity is low anyway, the estimate does not need to be as accurate as if the true sensitivity is high. On the other hand, we assume that estimating the sensitivity is more difficult if the true sensitivity is relatively low. Moreover, for the evaluation of the 'closeness' between s and an estimate for the true sensitivity \hat{s} , we ignore the direction of the deviation. Therefore, we define the weighted error of the point estimate as

$$weightederror(s, \hat{s}) := \frac{|s - \hat{s}|}{1 - s} \text{ for } s < 1, \quad (5)$$

where smaller values indicate better performance. For the ultimate decision whether the point estimate is close enough to the target value, we use an

indicator function which takes the value one if $weightederror(s, \hat{s})$ is below a given threshold value and zero otherwise.

Figure 3 illustrates the values of this indicator function combined with threshold values of 0.1, 0.3, and 0.5. The dark areas indicate combinations of true sensitivity and estimate of the true sensitivity that we consider successful with respect to the particular threshold value. For example, with a threshold value of 0.1, the estimate of the sensitivity has to be very close to the true value, even for low sensitivity values. In this study, we focus on a threshold value of 0.3. We will use this approach for the remainder of this study to evaluate the performance of the point estimate.

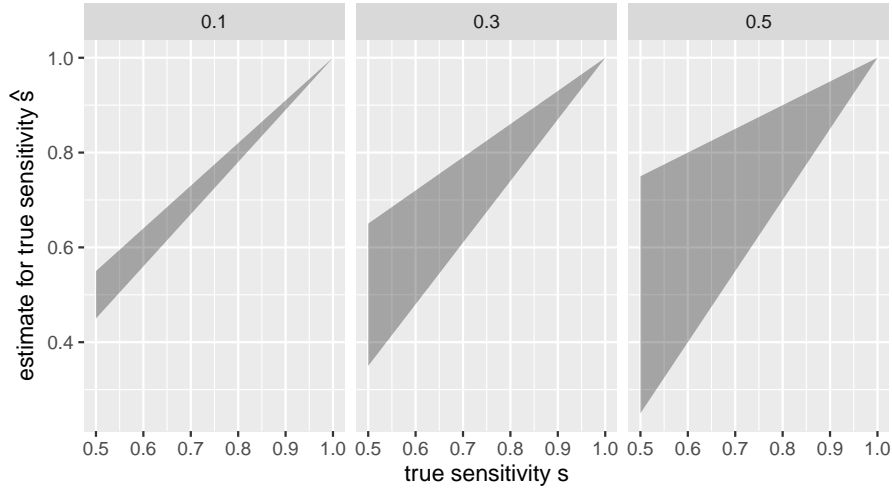


Figure 3: The dark areas in the panels above show successful combinations of true sensitivity s and estimate for the true sensitivity \hat{s} . Threshold values used from left to right: 0.1, 0.3, and 0.5.

Table 2 contains the same observations as the previous example but instead of the absolute error, the relative error combined with the threshold is used to derive the number of successful estimates among all runs. Although we observe approximately the same absolute error in all cases, the difference in the number of successful estimates between estimates for the degree centrality and estimates for the pagerank is up to nine percentage points. The (average) lower and upper values for a successful estimate are also shown in Table 2, i.e., if an estimate is within these boundaries, it is a successful estimate (boundaries and true sensitivity are averaged values for 500 trials).

Table 2: Example for two cases of the point estimate experiment. In both cases, we observe the same absolute error. But success rates for degree are lower.. (error mechanism: add edges 10%)

centrality measure	% successful imp.	% successful iter.	\hat{e}^{imp}	\hat{e}^{iter}	s
degree	0.938	0.910	0.018	0.018	0.942
pagerank	0.996	1.000	0.019	0.017	0.897

3.4 Results for synthetic graphs

Figure 4 illustrates the performance of the point estimates for the experiments on ER graphs. In general, the performance of the point estimate for the sensitivity in the context of Erdős–Rényi graphs is remarkable. The success rate, i.e. the fraction of cases where the point estimate is within the boundaries as defined in Section 3.3, is largely above 90%.

The success rates for all error mechanism and centrality measures are illustrated in Figure 4. In most cases, there is no difference between iterative and imputation estimate except for cases that involve the closeness centrality. In those cases, the imputation based estimate is better if edges are missing and for the error mechanism additional edges, the iterative estimate is better. This effect diminishes with increasing intensity.

The success rates for betweenness centrality, eigenvector centrality, and pagerank at the same, very high, level, followed by closeness and degree centrality. The latter two show slightly lower, but still high, success rates. The different error mechanisms show similar results. When comparing the 10% and 30% error mechanism, we cannot observe that the success rates are lower for the latter. The converse seems to be the case, the success rates for cases that involve error mechanisms with 30% error level are higher than the corresponding values for 10% error mechanisms. At first, this observation seems counter-intuitive. However, we also observe that the sensitivity decreases increasing measurement error. Since the sensitivity is lower, the interval for valid estimates becomes larger (Equation 5) and in the case of ER graphs, there is low variation and the success rates become better with increasing intensity.

We also perform the experiment described in Section 3.2 with one difference in the first step: instead of an ER graph, we generate a Barabási–Albert graph ([Barabási and Albert, 1999], 100 nodes, parameter $m = 11$, undirected). Results for this experiment for cases with 30% error level are shown in Figure 5. In general, the imputation based method performs than the iterative method. The performance of the imputation based method is particularly bad in cases where the betweenness centrality has to be estimated.

In contrast, the iterative method shows high success rates in most of the

cases. In cases that involve the degree or closeness centrality, the performance is usually worse than in the remaining cases. There is little difference between the four error mechanisms. The results for the 10% error mechanism (not shown) are similar. However, in some cases we observe lower success rates than in for the corresponding 30% error mechanism.

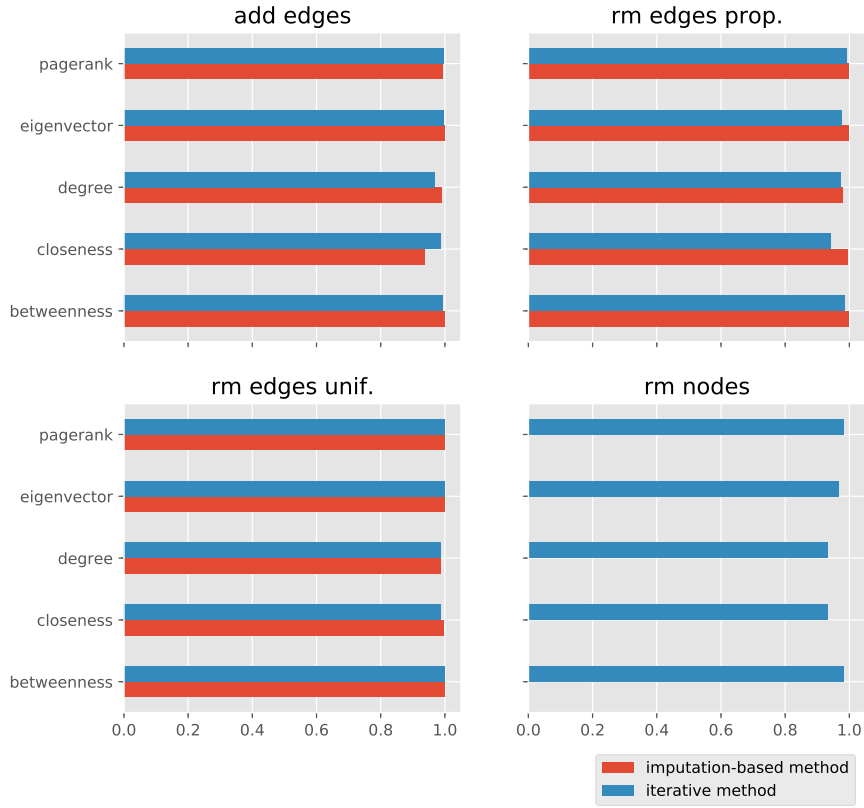


Figure 4: The the success rates for the point estimate (ER graphs, 30% error level) are shown in the figure above. The bar length indicates the percentage of successful cases among all trials (success rate). Red (blue) bars represent the success rates for imputation-based (iterative) estimation method.

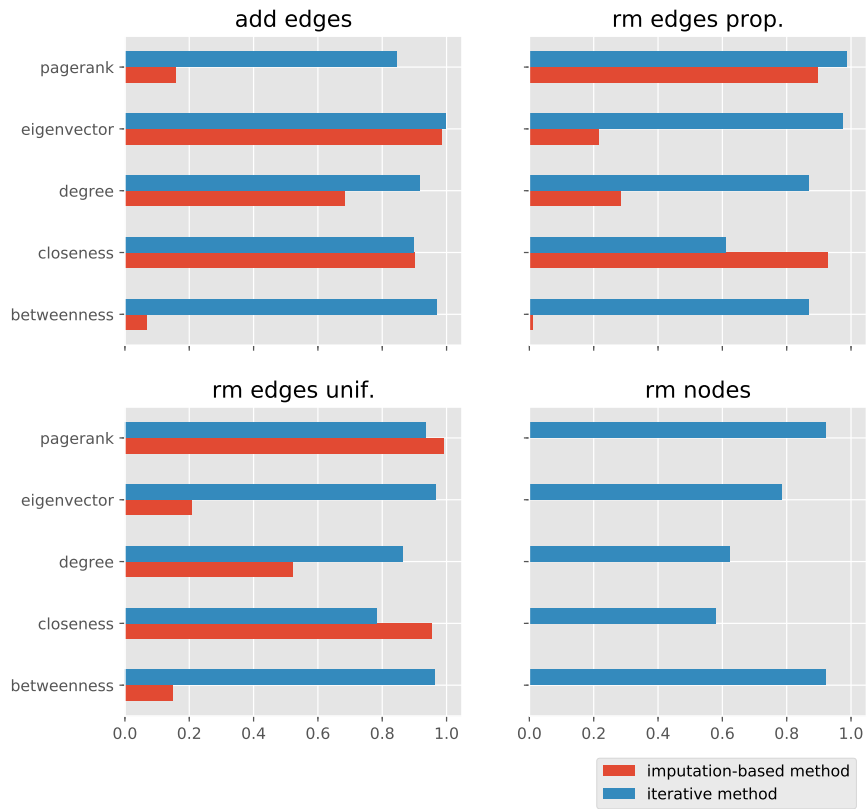


Figure 5: The the success rates for the point estimate (BA graphs, 30% error level) are shown in the figure above. The bar length indicates the percentage of successful cases among all trials (success rate). Red (blue) bars represent the success rates for imputation-based (iterative) estimation method.

4 Application to real-world networks

In this section, we apply our methods from Section 3.1 to real-world networks in order to investigate the suitability of these methods for practical application. We use four networks from different domains and thus different structural properties to get an impression how these methods perform on real data. Descriptive statistics for these networks [Kunegis, 2014] are listed in Table 3.

Table 3: Statistics of real-world networks. (If original network is not connected, we only consider the largest connected component)

Network	Nodes	Edges	Clustering	Density	Diameter
Dolphins	62	159	0.3029	0.0841	8
Jazz	198	2,742	0.6334	0.1406	6
Protein	1,458	1,948	0.1403	0.0018	19
Hamsterster	1,788	12,476	0.1655	0.0078	14

We use our proposed methods to estimate the sensitivity of five centrality measures under the influence of four error mechanisms. For every error mechanism, we consider two cases, a moderate scenario of 10% error level and a more intense scenario with 30% error level. For every combination of network, centrality measure, and error mechanism, the experimental setup is as follows:

1. Due to the very nature of the hidden networks, we can't access them. Hence, for the sake of our experiments, we treat the real-world network as the error-free hidden network H . (This is a common approach used in existing studies about the sensitivity of centrality measures.)
2. To simulate erroneous data collection, we choose a graph from $\varphi(H)$ and denote it by O . This graph represents the observed network which is affected by measurement errors. For evaluation purposes, the true sensitivity $\rho_c(H, O)$ is calculated and denoted by s .
3. Based on the observed network O , two estimates for the true sensitivity are calculated. The imputation based estimate ($\hat{\rho}_c^{imp}(O, H)$) is denoted by \hat{s}^{imp} , the estimate calculated according to the iterative method ($\hat{\rho}_c^{iter}(O, H)$) is denoted by \hat{s}^{iter} .

For every combination, we perform this experiment 500 times. To evaluate the results, we use the procedures described in Section 3.3.

4.1 Results for real-world networks

In this section, we study how our method for estimating the sensitivity of centrality measures performs on real-world networks. Regarding the iterative estimates, we observe a fair amount of cases with high success rate. However, the results for empirical networks are more heterogeneous than the results for Erdős–Rényi networks (Section 3.4). But since the real-world networks are more complex than the Erdős–Rényi networks, we expected that our estimation method wouldn’t work as well for real-world networks as for Erdős–Rényi networks.

The results for the estimate of the sensitivity of centrality measures in real-world networks are shown in Figure 6 and 7.

First, let’s focus on the results for error mechanisms with 10% error level (Figure 6). Comparing both estimation methods, it turns out that the iterative method is at least as good as the imputation method, except for a few cases where the former is slightly better. Hence we focus our discussion for the moment on the estimates of the iterative method. The iterative estimate for pagerank works in virtually all cases, regardless of the specific network or error mechanism. Among the four networks, the success rates for the Dolphins network are usually the lowest. It is reasonable to assume that this effect is due to the small size of the Dolphin network (62 nodes, 159 edges). If we focus our discussion on the three larger networks, we observe high success rates for the estimates of the sensitivity of closeness, betweenness, and eigenvector centrality if edges are missing uniformly or proportional. For these networks, the estimates for the sensitivity of the eigenvector centrality also work if the measurement errors lead to too many edges in the observed network. The results suggest, that the error mechanism missing nodes is most difficult one for the point estimate. There is no strong relation ship between sensitivity and the success rate, higher sensitivity values are not easier to estimate.

Comparing Figure 6 and 7 shows that the success rates for the iterative method become lower with increasing level of error. It is more difficult to estimate the sensitivity for higher levels of measurement errors. The estimate for the sensitivity of pagerank still works for error mechanisms additional edges and missing edges. If we focus on the three largest networks, the cases involving betweenness and closeness centrality show good success rates if edges are missing (uniformly and proportional). Cases involving the eigenvector centrality show good success rates if edges are missing uniformly and for the error mechanism additional edges.

We observe cases with a subpar performance for 10% error level where the success rates continue to decrease. For example, the Jazz network with additional edges in combination with the closeness and betweenness centrality. There are few cases that show good performance for 10% error level but work barely for 30% error level (e.g., if edges are missing proportionally

in the Jazz network and we try to estimate the sensitivity of eigenvector centrality or pagerank).

The results for the imputation based estimates of the sensitivity are rather different. There is no centrality measure or error mechanism where this method shows high success rates for all four networks. In some cases, the imputation estimate for error levels of 30% is better than the imputation based estimate for the corresponding 10% case (e.g., pagerank and additional edges). Most of these situations occur when the error mechanism is additional edges.

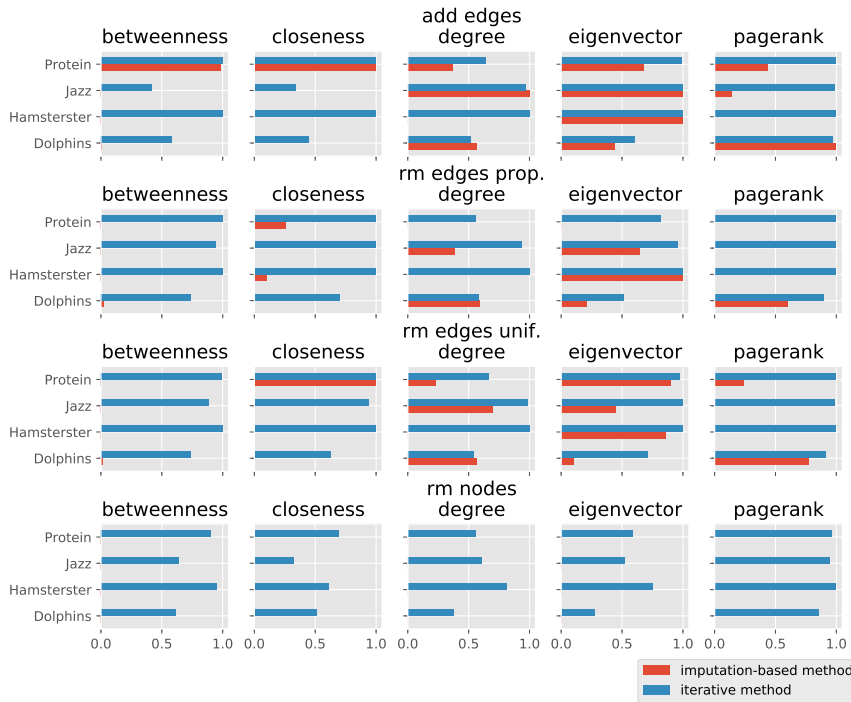


Figure 6: The success rates for the point estimate (real-world networks, 10% error level) are shown in the figure above. The bar length indicates the percentage of successful cases among all trials (success rate). Red (blue) bars represent the success rates for imputation-based (iterative) estimation method.

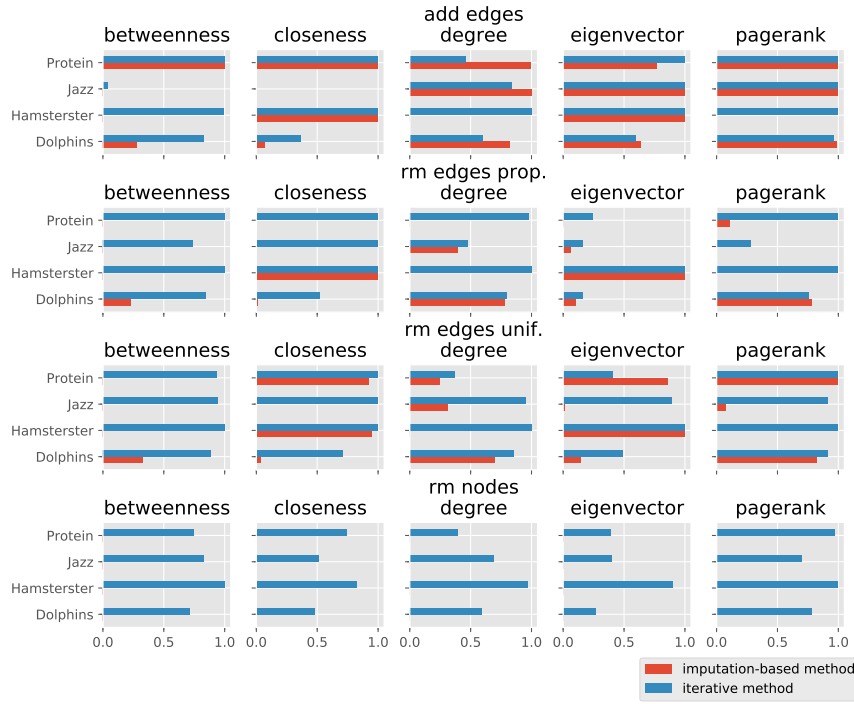


Figure 7: The success rates for the point estimate (real-world networks, 30% error level) are shown in the figure above. The bar length indicates the percentage of successful cases among all trials (success rate). Red (blue) bars represent the success rates for imputation-based (iterative) estimation method.

5 Discussion and conclusion

Errors in network data is a ubiquitous problem in network analysis and previous studies have shown that these errors can have severe impact on the reliability of centrality measures.

In this study we have introduced an easy to interpret definition for the sensitivity of centrality measures and a concept that models measurement errors as random graphs. We applied these concepts to ER graphs and find that the results are in general agreement with previous research ([Borgatti et al., 2006]).

We argued that the sensitivity has to be estimated and proposed two novel methods for estimation of the sensitivity: an imputation based approach and an iterative approach. Our experiments have shown that both methods perform very good on ER graphs. For BA graphs and real-world networks, the imputation based approach rarely works and should not be used. However, we could identify cases where the iterative estimation method shows good performance. It works especially well for the pagerank for all error mechanism with 10% error level. If the error level increases to 30%, it still shows good performance if edges are missing uniformly at random or if there are spurious edges. If 10% of the edges are missing uniformly at random or proportional and the network is not too small, the iterative method performs well for all centrality measures except for the degree centrality, though sensitivity values for the degree sensitivity are relatively high.

The results for the iterative estimation method are promising. It should be explored how we can infer from the observed data, whether this method should be used. Moreover, this estimation method is not limited to our definition of sensitivity. It would be interesting to see results for other metrics, for example estimation of the most central node (cf. [Frantz and Carley, 2016]).

References

- [Barabási and Albert, 1999] Barabási, A.-L. and Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, 286(October):509–512.
- [Bollobás and Riordan, 2002] Bollobás, B. and Riordan, O. (2002). Mathematical results on scale-free random graphs. *Handbook of Graphs and Networks: From the Genome to the Internet*, pages 1–38.
- [Bonacich, 1987] Bonacich, P. (1987). Power and Centrality: A Family of Measures. *American Journal of Sociology*, 92(5):1170–1182.
- [Borgatti et al., 2006] Borgatti, S. P., Carley, K. M., and Krackhardt, D. (2006). On the robustness of centrality measures under conditions of imperfect data. *Social Networks*, 28(2):124–136.

- [Brin and Page, 1998] Brin, S. and Page, L. (1998). The Anatomy of a Search Engine. *Proceedings of the 7th International World-Wide Web Conference*, page 20.
- [Butts, 2003] Butts, C. T. (2003). Network inference, error, and informant (in)accuracy: A Bayesian approach.
- [Costenbader and Valente, 2003] Costenbader, E. and Valente, T. W. (2003). The stability of centrality measures when networks are sampled. *Social Networks*, 25(4):283–307.
- [Csardi and Nepusz, 2006] Csardi, G. and Nepusz, T. (2006). The igraph Software Package for Complex Network Research. *InterJournal, Complex Systems*, (1695):1–9.
- [Frantz and Carley, 2016] Frantz, T. L. and Carley, K. M. (2016). Reporting a network’s most-central actor with a confidence level. *Computational and Mathematical Organization Theory*, pages 1–12.
- [Frantz et al., 2009] Frantz, T. L., Cataldo, M., and Carley, K. M. (2009). Robustness of centrality measures under uncertainty: Examining the role of network topology. *Computational and Mathematical Organization Theory*, 15(4):303–328.
- [Freeman, 1978] Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215–239.
- [Goodman and Kruskal, 1954] Goodman, L. A. and Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, 49(268):732–764.
- [Huisman, 2009] Huisman, M. (2009). Imputation of missing network data: some simple procedures. *Journal of Social Structure*, 10(1):1–29.
- [Kim and Leskovec, 2011] Kim, M. and Leskovec, J. (2011). The Network Completion Problem: Inferring Missing Nodes and Edges in Networks. *SIAM International Conference on Data Mining*, pages 47–58.
- [Koschützki et al., 2005] Koschützki, D., Lehmann, K., and Peeters, L. (2005). Centrality Indices. In Brandes, U. and Erlebach, T., editors, *Network Analysis: Methodological Foundations*, pages 16–61. Springer Berlin Heidelberg.
- [Kunegis, 2014] Kunegis, J. (2014). Handbook of Network Analysis [KONECT – the Koblenz Network Collection]. page 58.
- [Newman, 2017] Newman, M. E. J. (2017). Measurement errors in network data. *eprint arXiv:1703.07376*.

- [Smith and Moody, 2013] Smith, J. A. and Moody, J. (2013). Structural Effects of Network Sampling Coverage I: Nodes Missing at Random. *Social networks*, 35(4).
- [Smith et al., 2017] Smith, J. A., Moody, J., and Morgan, J. H. (2017). Network sampling coverage II: The effect of non-random missing data on network measurement. *Social Networks*, 48:78–99.
- [Wang et al., 2016] Wang, C., Butts, C. T., Hipp, J. R., Jose, R., and Lakon, C. M. (2016). Multiple imputation for missing edge data: A predictive evaluation method with application to Add Health. *Social Networks*, 45:89–98.
- [Wang et al., 2012] Wang, D. J., Shi, X., McFarland, D. A., and Leskovec, J. (2012). Measurement error in network data: A re-classification. *Social Networks*, 34(4):396–409.